

PATENT APPLICATION
TITLE: LATENCY MONITOR
INVENTOR: MICHAEL BUNDY
ATTORNEY DOCKET NUMBER: 30418US

5

SPECIFICATION

BACKGROUND

10

Broker-dealers increasingly make available to their customers on-line submission, cancellation, and tracking of the status of orders for securities. Securities trading customers include day traders, institutions, and active private investors. Such customers make many quick decisions regarding when and from whom to order securities.

15

Customers are increasingly demanding regarding quality and speed of execution. Customers require a high quality of information to support their decision-making. Customers are often presented with quotes identifying markets from which particular securities can be bought or sold at particular prices. In such quotes there are often several markets quoting securities at the current inside price. Markets quoting the same price, however, are not the same in terms of quality of execution. Especially regarding speed, all markets are different. It would be useful, therefore, if customers had a display of information helpful in identifying which markets are likely to execute orders more quickly than others.

20

25

Modern broker-dealers often subscribe to one or more exchanges or ECNs ("markets") capable of executing orders for securities by matching orders with orders of opposite side. Orders, cancellations, and responses are communicated to and from markets by use of data communications ports. Many broker-dealers handle volumes of orders so large as to require more than one port per market. Ports often are not equal in their ability to communicate with a particular market. Sometimes ports fail, partially or completely. It would be useful to have a display of information, for diagnostic purposes within the

30

00674595-054900

broker-dealer organizations, to help identify problems with particular ports, to help keep the overall flow of data communications functioning efficiently.

SUMMARY

5

One aspect of the invention provides methods for displaying latency. Embodiments of the invention are typically implemented in broker-dealer computer systems engaged generally in automated processing of orders for securities including sending to markets messages comprising orders and cancellations and receiving from markets responses to orders and cancellations. Embodiments include recording for messages sent to markets the time when each message is sent and the identity of the market to which each message is sent. Embodiments include recording for responses received from markets the time when each response is received, wherein each response corresponds to a particular message. Embodiments include calculating latencies for markets dependent upon recorded time when a message is sent to the market and a recorded time when a corresponding response is received from the market. Embodiments include latencies for ports as well as latencies for markets. Embodiments include displaying the identity of the markets and the latencies for the markets. Embodiments include counting and displaying the number of messages and responses received and sent during a period of time, for use in broker-dealer diagnostics.

A second aspect of the invention provides automated computing machinery, as system for calculating and displaying latency, typically implemented in broker-dealer computer systems capable of automated processing of orders for securities, includes sending messages to markets and receiving from markets responses to messages. Embodiments of this aspect include at least one computer processor programmed to record in computer memory, for messages sent to markets, the time when each message is sent and the identity of the market to which each message is sent. In such embodiments, processors are typically programmed also to record in computer memory, for responses received from markets, the time when each response is received. Each response corresponds to a particular message. In such embodiments, processors are programmed also to calculate

for markets latencies dependent upon recorded time when at least one message is sent to a market and recorded time when a corresponding response is received from the market. In such embodiments, processors typically are programmed also to display the identities of the markets and the latencies for the markets. Embodiments include latencies for ports as well as latencies for markets. Embodiments of this aspect typically include computer memory coupled to processors, the processors being further programmed to store in computer memory the latencies. Embodiments include processors programmed to count and display the number of messages and responses received and sent during a period of time, for use in broker-dealer diagnostics.

DRAWINGS

Figure 1A is a general data flow diagram showing various alternative embodiments of the invention.

Figure 1B is a detail of relations among ports and markets in various alternative embodiments of the invention.

Figure 2 illustrates calculating instant latency.

Figure 3 illustrates an alternative embodiment using average latency.

Figure 4 illustrates another alternative embodiment using average latency.

Figure 5A illustrates a form of display.

Figure 5B illustrates an alternative form of display.

Figure 5C illustrates an alternative form of display.

Figure 6 illustrates computing machinery for various alternative embodiments of the invention.

Figure 7 illustrates computing machinery programmed to calculate instant latency.

Figure 8 illustrates computing machinery programmed to count and display the number of messages and responses received and sent during a period of time.

Figure 9 illustrates computing machinery programmed to count and display the number of messages and responses received and sent through a port during a period of time.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

Definitions:

- 5 “Cancellation” is termination of an order, or partial termination of an order, by the customer or by software comprising an embodiment of the invention. In addition, markets can cancel orders, or parts of orders, for example, in response to an IOC order.

- “ECN” abbreviates “Electronic Communications Network,” referring to an order
10 matching service that provides liquidity by matching orders rather than by maintaining inventory. In the context of the invention, ECNs are considered markets. ECNs, like market makers are identified by use of market participant identification codes or “MPIDs.” In order to avoid confusion with data communications networks, ECNs are referred to as either “ECNs” or as “markets.” Some current ECNs, their symbols and
15 names, are listed below. The number and identities of ECNs changes from time to time.

<u>Example List of ECNs</u>	
<u>MPID</u>	<u>Name</u>
ARCA	Archipelago
BTRD	Bloomberg Trade Book
INCA	Instinet
ISLD	Island
MWSE	Midwest Stock Exchange
NTRD	NexTrade
REDI	Speer Leeds

“Exchange” means a national, regional, or international exchange for securities trading including for example, Nasdaq or NYSE.

5 “Executed,” in reference to an order, means that shares have been either bought or sold according to the side of the order.

10 “Filled” means executed. That is, all shares in the order have been executed, bought or sold according to the side of the order. If an order is subject to partial fulfillment, then the order can be partly filled and partly rejected or cancelled, in which case the order will never be considered filled. Processing of an order can therefore be completed through some combination of cancellation, rejection, killing, and partial execution without the order’s ever being filled. Processing of an order is said to be complete when all the shares in the order, share by share, have been executed, cancelled, rejected, or killed.

15 “Inside price” means, as appropriate, the highest bid price or the lowest ask price for a particular security. For buy orders, the inside price is the lowest ask price. For sell orders, the inside price is the highest bid price.

20 “Latency” means a measure of the speed with which markets respond to orders and cancellations. Latency in many embodiments of the invention is determined as the difference between the time when a response to an order is received and the time when the corresponding order was routed to the market. Latency can be measured from normal orders or from test orders. Some markets support test orders as such. For markets in which test orders as such are not supported, test orders can be implemented by use of unmarketable orders immediately followed by cancellations. For markets receiving orders regularly, latency can be tracked from normal orders, without the need for test orders. Latency can be embodied as a single ratio difference between two recorded times or as various kinds of averages.

30 “Level Two Quotes” are quotes that comprise one or more market participant quotes (“MPQs”). The best known source of level two quotes is Nasdaq, but “level two quotes”

refers to any form of market information that aggregates market participant quotes for a security.

“Market,” “electronic market,” “market participant,” “electronic market participant,” “marketing network,” and electronic marketing network” are all used as synonyms for services accessible through electronic communications networks capable of executing orders for securities by accepting from broker-dealers buy orders and sell orders, matching or failing to match buy orders with sell orders, and communicating the results to the broker-dealers. Generally the term “market” is used to refer to these entities. All “markets,” as the term is used, are either ECNs or market makers. All available markets have names and symbols as described under the definitions of “ECN” and “market maker.”

“Market maker” means a broker-dealer providing order matching and liquidity in a stock by maintaining an inventory of the stock. Market makers typically trade their inventories through exchanges. Some currently active market makers, their symbols and names, are listed below. The number and identity of market makers can change from time to time.

Example List of Market Makers

<u>MPID</u>	<u>Name</u>
BEST	Bear, Stearns & Co., Inc.
BTAB	Alex, Brown & Sons, Inc.
GSCO	Goldman, Sachs & Co.
HMQT	Hambrecht & Quist, LLC
HRZG	Herzog, Heine, Geduld, Inc.
JANY	Janney Montgomery Scott, Inc.
LEHM	Lehman Brothers, Inc.

MADF	Bernard L. Madoff
MLCO	Merrill Lynch, Pierce, Fenner & Smith Inc.
MOKE	Morgan, Keehan & Co., Inc.
MONT	Nationsbanc Montgomery Securities, LLC
MSCO	Morgan Stanley & Co., Inc.
NITE	Knight Securities, L.P.
OLDE	Olde Discount Corporation
OPCO	CIBC Oppenheimer Corporation
PIPR	Piper Jaffray Inc.
PRUS	Prudential Securities, Inc.
PWJC	Paine Webber, Inc.
RAJA	Raymond James & Associates, Inc.
SBSH	Smith Barney, Inc.
SHRP	Sharpe Capital, Inc.
SHWD	Sherwood Securities Corporation

“Orders” are orders for purchase or sale of securities. In many of the embodiments described, “orders” are electronic orders for purchase or sale of securities.

- 5 “Quotes” are aggregates of information regarding securities traded in markets. Quotes include for securities listed for sale or purchase, symbols identifying the securities, price, side, quantities, and market identifications or MPIDs. Quotes can come from exchanges or directly from markets. A “Nasdaq Level Two Quote” includes market information in the form of market participant quotes for all markets offering to buy or sell a particular
- 10 security through Nasdaq.

“Securities” are any agreement for investment. Stocks are the securities most often addressed in described embodiments of the invention. The invention, however, is

applicable to many kinds of securities including, for example, options, commodities, and bonds.

“Side” refers to which side of the market is represented by an order or a quote. Side indicates whether the quote or order is to buy or sell, bid or ask. “Bid” indicates the buy side. “Ask” indicates the sell side. The present invention functions equally for either side of a transaction. Therefore we attempt to speak in neutral terms regarding side. We speak of execution rather than buying or selling. We use the term “price improvement” to indicate both price reductions for buy orders and price increases for sell orders.

Detailed Description:

Turning now to Figure 1A, a first aspect of the invention is seen. One embodiment illustrated in Figure 1A provides a method of displaying latency. The embodiment is implemented in a broker-dealer computer system. The system is engaged in automated processing of orders (150) for securities including sending (104) messages (102) to markets (108) and receiving (112) from markets (108) responses (114) to messages.

The illustrated embodiment includes recording (106) for messages sent to markets the time (120) when each message is sent and the identity (118) of the market to which each message is sent, the messages (102) comprising orders (150) and cancellations (148) of orders. This first embodiment includes also recording (110) for responses received from markets the time (122) when each response is received, wherein each response corresponds to a particular message.

This embodiment includes also calculating (124) for at least one market a latency (128) dependent upon at least one recorded time (120) when at least one message is sent to the market and at least one recorded time (122) when a corresponding response is received from the market. The illustrated embodiment includes displaying (130) the identity (118) of the market and the latency (128) for the market. In a further embodiment shown in

Figure 1A, latency (128) is a latency for a port (154), the port being identified by Port ID code (156).

Shown in Figure 5A is an example of a form of display useful with many embodiments of the invention. The example in Figure 5A illustrates a columnar display of identities (118) of markets and latency implemented as an instant latency (502) for each market and an average latency (504) for each market.

As shown on Figure 1A, the display function (130) in many embodiments sends (133) the display (135) to display devices (134) by use of data communications (132). Data communications (132) in some embodiments includes networks, such as intranets, extranets, or internets, and in other embodiments includes satellite channels, direct telephone links, and other forms of data communications. Use of any form of data communications is well within the invention.

In a further embodiment, shown in Figure 2, latency is implemented as an instant latency (202). The instant latency (202) is calculated (204) dependent upon one recorded time (120) when one message is sent to a market and one recorded time (122) when a corresponding response is received from the market.

In a still further embodiment, shown in Figure 3, latency is implemented as an average latency (320). The average latency (320) is dependent upon at least one recorded time (306, 314) when at least one message is sent to the market and at least one recorded time (308, 316) when a corresponding response is received from the market. In embodiments of the kind shown in Figure 3, the recorded times (306, 308, 314, 316) used in calculating the average latency (320) are recorded during a defined period of time (322).

In a further embodiment, shown in Figure 4, the latency is implemented as an average latency (420). The average latency is dependent upon at least one recorded time (408, 414) when at least one message is sent to the market and at least one recorded time (410, 416) when a corresponding response is received from the market. In embodiments shown

in Figure 4, the number of recorded times (408, 410, 414, 416) used to calculate the average latency (420) is limited to a defined maximum number "N" (422). In many embodiments of this kind, the N recorded times used to calculate average latency are the N most recent recorded times.

A further embodiment shown in Figure 1A includes the steps of counting (136) the number of messages sent to at least one market during a period of time, including storing in computer memory (140) the number of messages (144) sent to the market during the period of time. In many embodiments of this kind, the counting steps (138, 136) determine time periods in dependence upon a computer system clock (170).

The illustrated embodiment includes also counting (138) the number of responses received from the market during the period of time, including storing in computer memory (140) the number of responses (146) received from the market during the period of time. The embodiment includes also displaying (130), in addition to the identity (118) of the market and the latency (128) for the market, the number of messages (144) sent to the market and the number of responses (146) received from the market during the period of time.

Figure 5C shows an example of a display useful with various embodiments using such displays display in columnar form the market identities (118), latencies (320), number of messages sent during a period of time (144), and the number of responses received during a period of time (146). As shown on Figure 1A, the display function (130) in many embodiments sends (133) the display (135) to display devices (134) by use of data communications (132). Data communications (132) in some embodiments includes networks, such as intranets, extranets, or internets, and in other embodiments includes satellite channels, direct telephone links, and other forms of data communications. Use of any form of data communications is well within the invention.

An example of the use of message counts for diagnostic purposes is a display showing an increase in latency for a port explained by an increase in message counts for the port, thus

indicating the port slowed down because its work load increased, and indicating also that there is no problem with the system. Another example is a display showing an increase in latency for a port explained by the port's message count going to zero, thus indicating that the increase in latency is caused by a catastrophic failure of the port.

5

In a further embodiment, shown in Figure 1A, the system includes the steps of counting (136) the number of messages sent to a market through a port (154) during a period of time, including storing in computer memory (140) the number of messages (144) sent to the market through the port during the period of time. In many embodiments of this kind,
10 the counting steps (138, 136) determine time periods in dependence upon a computer system clock (170).

The illustrated embodiment includes also counting (138) the number of responses received from the market through the port during the period of time, including storing in
15 computer memory (140) the number of responses (146) received from the market through the port during the period of time. The system includes also displaying (130), in addition to the identity (118) of the market and the latency (128) for the market, the number of messages (144) sent to the market through the port and the number of responses (146) received from the market through the port during the period of time.

20

Figure 5B shows an example of a display useful with various embodiments of the invention. Embodiments using such displays display in columnar form the market identities (118), port identity codes (154), instant latencies (202), average latencies (320), number of messages sent during a period of time (144), and the number of responses
25 received during a period of time (146).

As shown on Figure 1A, the display function (130) in many embodiments sends (133) the display (135) to display devices (134) by use of data communications (132). Data communications (132) in some embodiments includes networks, such as intranets,
30 extranets, or internets, and in other embodiments includes satellite channels, direct

telephone links, and other forms of data communications. Use of any form of data communications is well within the invention.

Turning now to Figure 6, an additional aspect of the invention is seen. One embodiment shown in Figure 6 is automated computing machinery implementing a broker-dealer computer system (602). The illustrated embodiment is capable of automated processing of orders for securities, including sending (606) messages (604) to markets (608) and receiving (610) from markets (608) responses (612) to messages.

The illustrated embodiment includes at least one computer processor (618) programmed to record (622) in computer memory (620), for messages sent to markets, the time (632) when each message is sent and the identity (630) of the market (608) to which each message is sent, the messages including orders (614) and cancellations (616) of orders. In this embodiment, the processor is programmed also to record (624), in computer memory (620), for responses (612) received (610) from markets, the time (634) when each response is received. Each response (612) corresponds (642) to a particular message (604).

In this example embodiment, the processor is programmed also to calculate (626), for at least one market (608) a latency (628) dependent upon at least one recorded time (632) when at least one message is sent to the market and at least one recorded time (634) when a corresponding response is received from the market. In this embodiment, the processor is programmed also to display (632) of the identity (630) of the market and the latency (628) for the market. In a further embodiment shown in Figure 6, latency (628) for a market (608) is also latency for a port (644), the port being identified in data by a port ID code (646).

As shown on Figure 6, the display function (632) in many embodiments sends (633) the display (635) to display devices (638) by use of data communications (636). Data communications (636) in some embodiments includes networks, such as intranets, extranets, or internets, and in other embodiments includes satellite channels, direct

telephone links, and other forms of data communications. Use of any form of data communications is well within the invention. The embodiment illustrated in Figure 6 includes also computer memory (620) coupled (640) to the processor (618), the processor being further programmed to store (630) in computer memory (620) the latency (628).

5

In a further embodiment shown in Figure 7, latency comprises an instant latency. In the embodiment of Figure 7, the processor (618) is programmed to calculate (626) latency as an instant latency (702) calculated dependent upon one recorded time (632) when one message is sent to a market and one recorded time (634) when a corresponding response is received from the market.

10

In some embodiments, the processor is programmed to calculate latency as an average latency dependent upon at least one recorded time when at least one message is sent to the market and at least one recorded time when a corresponding response is received from the market. In such embodiments, as illustrated in Figure 3, recorded times (306, 308, 314, 316) used in calculating the average latency (320) are recorded during a defined period of time (322).

15

In other embodiments, the processor is programmed to calculate latency as an average latency dependent upon at least one recorded time when at least one message is sent to the market and at least one recorded time when a corresponding response is received from the market. In such embodiments, as shown in Figure 4, the number of the recorded times (408, 410, 414, 416) used to calculate the average latency is limited to a defined maximum number "N" (422). In many embodiments of this kind, the N recorded times used to calculate average latency are the N most recent recorded times.

20

25

In a further embodiment shown in Figure 8, latency comprises an average latency. In embodiments of the kind shown in Figure 8, the processor (618) is further programmed to count (802) the number (806) of messages (604) sent (606) to at least one market (608) during a period of time, including storing in computer memory (620) the number of

30

messages (806) sent to the market during the period of time. Periods of time in many embodiments are determined dependent upon a system clock (816).

The processor in many embodiments of the kind illustrated in Figure 8, is also
5 programmed to count (804) the number (808) of responses (612) received (610) from the market (608) during the period of time, including storing in computer memory (620) the number of responses (808) received from the market during the period of time. The system includes also displaying (814), in addition to the identity of the market and the latency for the market, the number of messages (806) sent to the market and the number
10 of responses (808) received from the market during the period of time.

Figure 5C shows an example of a display useful with various embodiments of the invention. Embodiments using such displays display in columnar form the market identities (118), latencies (320), number of messages sent during a period of time (144),
15 and the number of responses received during a period of time (146). As shown on Figure 8, the display function (814) in many embodiments sends (815) the display (817) to display devices (638) by use of data communications (636). Data communications (636) in some embodiments includes networks, such as intranets, extranets, or internets, and in other embodiments includes satellite channels, direct telephone links, and other forms of
20 data communications. Use of any form of data communications is well within the invention.

In a further embodiment shown in Figure 9, the processor (618) is programmed to count (904) the number of messages (604) sent to a market (608) through a port (902) during a
25 period of time, including storing (914) in computer memory (620) the number of messages (912) sent to the market (608) through the port (902) during the period of time. Periods of time in such embodiments typically are determined dependent upon a system clock (816).

30 The processor in many embodiments of the kind illustrated in Figure 9 is also programmed to count (906) the number of responses (612) received (610) from the

market (608) through the port (902) during the period of time, including storing (916) in computer memory (620) the number of responses (910) received from the market (608) through the port (902) during the period of time. The embodiment as illustrated includes also displaying (908), in addition to the identity (630 on Figure 6) of the market and the latency (628 on Figure 6) of the market, the number of messages (912) sent to the market (608) through the port (902) and the number of responses (910) received from the market (608) through the port (902) during the period of time.

Figure 5B shows an example of a display useful with various embodiments of the invention. Embodiments using such displays display in columnar form the market identities (118), port identity codes (154), instant latencies (202), average latencies (320), number of messages sent during a period of time (144), and the number of responses received during a period of time (146). As shown on Figure 9, the display function (908) in many embodiments sends (909) the display (911) to display devices (638) by use of data communications (636). Data communications (636) in some embodiments includes networks, such as intranets, extranets, or internets, and in other embodiments includes satellite channels, direct telephone links, and other forms of data communications. Use of any form of data communications is well within the invention.